

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2003, 14:17:59 ; Search time 2795 Seconds
(without alignments)
6410.882 Million cell updates/sec

Title: US-09-787-216A-2

Perfect score: 438
Sequence: 1 gtcacctgactcttgaggagga.....ccctggcccaagatcatcac 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sis.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vi.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_man.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	428.4	97.8	438	6	A01592 Human haemo
2	428.4	97.8	438	6	AR050456 Sequence
3	428.4	97.8	438	6	AR124584 Sequence
4	428.4	97.8	438	6	AR142760 Sequence
5	428.4	97.8	438	6	I27203 Sequence 4
6	428.4	97.8	438	6	I63086 Sequence 4
7	428.8	97.4	438	6	A78805 Sequence 3
8	428.8	97.4	438	6	AF117710 Homo sapi
9	428.8	97.4	438	6	BC007075 Homo sapi
10	428.8	97.4	438	6	AR030837 Sequence
11	425.2	97.1	534	9	AY136510 Homo sapi
12	425.2	97.1	579	9	AF181989 Homo sapi
13	425.2	97.1	581	9	AF349114 Homo sapi
14	425.2	97.1	626	9	AF349114 Homo sapi
15	423.6	96.7	444	6	AR062963 Sequence
16	423.6	96.7	544	6	A18171 Beta globin
17	423.6	96.7	743	6	AR228705 Sequence
18	422	96.3	549	9	HSBGLX
19	422	96.3	1140	6	AR257263 Sequence
20	422	96.3	1140	6	AR257265 Sequence
21	421.4	96.2	576	9	HUMHEMOB
22	420.4	96.0	743	6	AR032231 Sequence
23	420.4	96.0	743	6	BD011603 Recombina
24	420.4	96.0	743	6	BD165862 Transgeni
25	420.4	96.0	743	6	I32884 Sequence 9
26	393.2	83.8	468	6	AX336431 Sequence
27	393.2	83.8	468	6	AX409658 Sequence
28	393.2	83.8	468	6	AX409658 Sequence
29	377.8	86.3	790	9	HUNBETGLA
30	376.8	85.0	475	6	AX400061 Sequence
31	376	85.8	488	9	AF181832 Homo sapi
32	374.6	85.5	438	6	AR050454 Sequence
33	374.6	85.5	438	6	AR124582 Sequence
34	353.2	80.6	589	4	OCBGL2
35	353.2	80.6	589	6	AX419806 Sequence
36	353.2	80.6	589	6	BD025503 Sequence
37	350.4	80.0	454	6	AX448896 Sequence
38	347.2	79.3	498	6	AX448896 Sequence
39	340.4	77.7	438	4	MIRRNAB
40	310	70.8	577	10	MI7084 Rat major b
41	310	70.8	620	10	X16417 Rat mRNA fo
42	305.2	69.7	441	10	AB020013 Mus muscu
43	305.2	69.7	441	10	AB020015 Mus muscu
44	305.2	69.7	441	10	AB020017 Mus muscu
45	305.2	69.7	444	10	AF149782 Mus muscu

ALIGNMENTS

RESULT 1	A01592	438 bp	DNA	linear	PAT 08-MAR-1993
LOCUS	Human haemoglobin A beta chain.				
DEFINITION	A01592				
ACCESSION	A01592				
VERSION	A01592.1	GI:344421			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 438)				
JOURNAL	BLOOD SUBSTITUTES				
	Patent: WO 8809179-A 10 01-DEC-1988;				

Pred. No. is the number of results predicted by chance to have a

```

FEATURES             Location/Qualifiers
     source            1..438
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
     CDS               <1_>438
                        /codon_start=1
                        /product="haemoglobin A beta chain"
                        /protein_id="CAA00182.1"
                        /db_xref="GI:344422"
                        /translation="VHLTPREKSAVTAALMGKVNVDVGEALGRLIVVWPTORFPES
                        FGLSTPDVANGNPKVAKHGKVLGAFSDGLAHLNLDKGFATLSLHCDKLHVDPEN
                        FRLLGNVLVCVLAHFHKEFTFPVQAAQKVWAGVANALAHKVH"
BASE COUNT           85 a 114 c 135 g 104 t
ORIGIN
1 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
2 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
361 GAATTCACCCCAAGTATATCAC 438
361 GAATTCACCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438

RESULT 3
LOCUS               AR124584
DEFINITION          Sequence 6 from patent US 6172039.
ACCESSION            AR124584
VERSION              AR124584.1
KEYWORDS              GI:14109945
SOURCE               Unknown.
ORGANISM             Unclassified.
REFERENCE            1 (bases 1 to 438)
AUTHORS              De Angelo, J., Motwani, N.M., Bajwa, W. and Bonaventura, J.
TITLE                Expression of recombinant hemoglobin and hemoglobin variants in
                    Yeast
JOURNAL              Patent: US 6172039-A 6 09-JAN-2001;
FEATURES             Location/Qualifiers
     source            1..438
                        /organism="unknown"
BASE COUNT           85 a 114 c 135 g 104 t
ORIGIN
1 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
2 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
361 GAATTCACCCCAAGTATATCAC 438
361 GAATTCACCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438

RESULT 2
LOCUS               AR050456
DEFINITION          Sequence 6 from patent US 5827693.
ACCESSION            AR050456
VERSION              AR050456.1
KEYWORDS              GI:5973181
SOURCE               Unknown.
ORGANISM             Unclassified.
REFERENCE            1 (bases 1 to 438)
AUTHORS              De Angelo, J., Motwani, N.M., Bajwa, W. and Bonaventura, J.
TITLE                Expression of recombinant hemoglobin and hemoglobin variants in
                    Yeast
JOURNAL              Patent: US 5827693-A 6 27-OCT-1998;
FEATURES             Location/Qualifiers
     source            1..438
                        /organism="unknown"
BASE COUNT           85 a 114 c 135 g 104 t
ORIGIN
1 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
2 GTGACCTGACTCTCTGAGGAGAGTGGCCCTTACTGCGCTGTGGGCAAGGTGAACGTG 60
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
61 GATGAAGTTGGTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 120
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
121 TTCTTTGAGTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCCCTAAGGTG 180
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
181 AAGGCTCATGGCAAGAAAGTGTCTGGTGCCTTTAGTATGATGGCTGCTCACTGGACAAC 240
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
241 CTCAAGGGCACTTTGGCACTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 GAGAACTTCAAGTCTCTGGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
361 GAATTCACCCCAAGTATATCAC 438
361 GAATTCACCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438
421 CTGCCCCCAAGTATATCAC 438

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2003, 14:24:09 ; Search time 2635 Seconds
(without alignments)
4039.985 Million cell updates/sec

Title: US-09-787-216A-2
Perfect score: 438
Sequence: 1 ggcacactgactctgagga.....ccctggccacaaagtatcac 438

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 4562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rtd:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428.4	97.8	508	12 BM263008	BM263008 ig25c03.y
2	428.4	97.8	515	13 BU658312	BU658312 cl34f02.z
3	428.4	97.8	522	13 BU660954	BU660954 cl66cl2.z
4	428.4	97.8	566	12 BM690339	BM690339 UI-E-ClD-

C	5	428.4	97.8	572	12 BM142584	BM142584 ig25c03.x
	6	428.4	97.8	573	12 BG943930	BG943930 ax43h03.x
	7	428.4	97.8	580	13 BU665464	BU665464 cl131e08.
	8	428.4	97.8	588	12 BI909771	BI909771 603070856
C	9	428.4	97.8	590	12 BM091317	BM091317 ig21b12.x
C	10	428.4	97.8	591	12 BM090889	BM090889 ig17a08.x
C	11	428.4	97.8	591	12 BQ029722	BQ029722 UI-H-DT0-
C	12	428.4	97.8	593	13 BU664237	BU664237 cl113e05.
C	13	428.4	97.8	595	12 BM994174	BM994174 UI-H-DT0-
	14	428.4	97.8	598	13 BU661951	BU661951 cl79a10.z
	15	428.4	97.8	598	13 CB270453	CB270453 1009360.H
	16	428.4	97.8	601	10 BG334450	BG334450 602553685
	17	428.4	97.8	601	10 AW950603	AW950603 EST362673
	18	428.4	97.8	601	14 CD107933	CD107933 AGENCOURT
	19	428.4	97.8	603	9 AV722370	AV722370 AV722370
C	20	428.4	97.8	605	12 BM662991	BM662991 UI-B-ClD-
C	21	428.4	97.8	606	12 BM992829	BM992829 UI-H-DT0-
C	22	428.4	97.8	607	13 BQ47183	BQ47183 UI-H-EU1-
C	23	428.4	97.8	608	13 BU683550	BU683550 UI-CF-ECL
C	24	428.4	97.8	608	14 CA427861	CA427861 UI-H-DFO-
C	25	428.4	97.8	609	12 BM597046	BM597046 170006874
C	26	428.4	97.8	609	13 BU682597	BU682597 UI-CF-ECL
	27	428.4	97.8	611	13 BU073853	BU073853 in25d09.Y
	28	428.4	97.8	615	12 BM613468	BM613468 170006592
C	29	428.4	97.8	617	13 BU617711	BU617711 UI-H-DFO-
C	30	428.4	97.8	619	12 BM993095	BM993095 UI-H-DT0-
C	31	428.4	97.8	623	14 CA435390	CA435390 UI-H-DT0-
	32	428.4	97.8	626	13 BU662685	BU662685 cl89e10.z
	33	428.4	97.8	628	14 CD559383	CD559383 AGENCOURT
C	34	428.4	97.8	628	12 BM992554	BM992554 UI-H-DT0-
	35	428.4	97.8	628	13 BU662291	BU662291 cl84a07.z
	36	428.4	97.8	628	14 CD108766	CD108766 AGENCOURT
C	37	428.4	97.8	629	12 BM992721	BM992721 UI-H-DT0-
C	38	428.4	97.8	629	13 BU617931	BU617931 UI-H-DFO-
	39	428.4	97.8	629	14 CD522665	CD522665 AGENCOURT
C	40	428.4	97.8	630	12 BM664514	BM664514 UI-B-ClD-
C	41	428.4	97.8	630	13 BU626306	BU626306 UI-H-DFO-
	42	428.4	97.8	630	14 CD520923	CD520923 AGENCOURT
C	43	428.4	97.8	630	14 CD559221	CD559221 AGENCOURT
C	44	428.4	97.8	631	14 CA427969	CA427969 UI-H-DFO-
C	45	428.4	97.8	632	14 CA425145	CA425145 UI-H-DFO-

ALIGNMENTS

RESULT 1

BM263008

LOCUS

DEFINITION

ig25c03.y1 Human Fetal Pancreas 1B Homo sapiens cdna 5', similar to

SW:HBB_HUMAN P02023 HEMOGLOBIN BETA CHAIN. [3] SW:HBB_HUMAN

TR:Q14510 ; mRNA sequence.

ACCESSION

BM263008

VERSION

BM263008.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 508)

AUTHORS

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,

Lemishka, I., Seearce, M., Brstelli, J., Gradwohl, G., Clifton, S.,

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blstein, A.,

Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas

, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.

, Jackson, Y. and Bowers, Y.

Endocrine Pancreas Consortium

Unpublished

Other ESTs: ig25c03.x1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmeltan@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@imgate.wustl.edu)

Putative full length read

vector to vector length is

Seg primer: -40RP from Gibco

High quality sequence stop: 476.

Location/Qualifiers

FEATURES

source

1. .508

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="Fetal pancreas (4 Pooled Donors, 18 - 20

weeks, Stratagene #738023)"

/clone_lib="Human Fetal Pancreas 1B"

/dev_stage="Fetal Pancreas"

xhoI: cDNA made by oligo-dT priming. Size-selected on

agarose gel. Average insert size ~1kb. 5' XhoI site was

destroyed after directional cloning. Amplified once.

Contact information: Hiroshi Inoue, MD, Metabolism Div.

(Alan Permutt Lab), Washington University School of

Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO

63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,

Fax: 314-747-2692."

96 a 135 c 147 g 130 t

BASE COUNT

ORIGIN

Query Match 97.8%; Score 428.4; DB 12; Length 508;

Best Local Similarity 98.6%; Pred. No. 5.2e-103;

Matches 432; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GTGCACCTGACTCTCTGAGAGAGTGGCGGCTTACTGCGCTGTGGGCAAGTGAACGTG 60

Db 15 GTGCACCTGACTCTCTGAGAGAGTGGCGGCTTACTGCGCTGTGGGCAAGTGAACGTG 74

QY 61 GATGAAGTTGGTGGTGGAGCTGGGCGGCTGGTGGTGTCTACCTTGGACCCAGAGG 120

Db 75 GATGAAGTTGGTGGTGGAGCTGGGCGGCTGGTGGTGTCTACCTTGGACCCAGAGG 134

QY 121 TTCTTTGAGTCTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCTTAAGGTG 180

Db 135 TTCTTTGAGTCTCTTTGGGATCTGTCCACTCTCTGATGCTGTTATGGGCAACCTTAAGGTG 194

QY 181 AAGGCTCATGGCAAGAAAGTGTCTGGTGGCTTTAGTATGGCTGGCTCACTGACACAC 240

Db 195 AAGGCTCATGGCAAGAAAGTGTCTGGTGGCTTTAGTATGGCTGGCTCACTGACACAC 254

QY 241 CTCAGGGCACCTTTTCCACACTGATGAGTGCATGCTGACAGAGTGCACGTGATCCT 300

Db 255 CTCAGGGCACCTTTTCCACACTGATGAGTGCATGCTGACAGAGTGCACGTGATCCT 314

QY 301 GAGAACTTCAGGCTCTCTGGGCAACGTGCTGGTGGTGTCTGCGCCCATCACTTGGCAA 360

Db 315 GAGAACTTCAGGCTCTCTGGGCAACGTGCTGGTGGTGTCTGCGCCCATCACTTGGCAA 374

QY 361 GAATTACCCACCACTGACAGGCTGCTATCAGAAAGTGGTGGCTGATGCTATGCC 420

Db 375 GAATTACCCACCACTGACAGGCTGCTATCAGAAAGTGGTGGCTGATGCTATGCC 434

QY 421 CTGGCCCAACAAGTATCAC 438

Db 435 CTGGCCCAACAAGTATCAC 452

RESULT 2

BU658312

LOCUS

DEFINITION

cl134f02.z1 Hembase; Erythroid Precursor Cells (LCB:cl library) Homo

ACCESSION

BU658312

VERSION

BU658312.1 GI:23370494

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (Bases 1 to 515)

AUTHORS

Gubin,A.N., Lee,Y.T., Bouffard,G.G. and Miller,J.L.

TITLE

Gene Expression in Human Erythroid Precursor Cells

JOURNAL

Unpublished

COMMENT

Contact: Jeffery L. Miller

Laboratory of Chemical Biology

National Institute of Diabetes and Digestive and Kidney Diseases

Building 10, Room 9B17, National Institutes of Health, Bethesda, MD

20892, USA

Tel: 301 402 2373

Fax: 301 435 5148

Email: jlm7@nih.gov

The 'cl' library was constructed by Alexander Gubin, Ph.D. in the

Laboratory of Chemical Biology, NIDDK, NIH. DNA Sequencing and/or

analyses by National Institutes of Health Intramural Sequencing

Center (NISC). More information available at:

http://hembase.nidk.nih.gov

Plate: 34 row: f column: 02

Seq primer: 5' lambda-TripEx2 Sequencing Primer.

FEATURES

source

1. .515

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="cl134f02"

/sex="unknown"

/tissue_type="blood"

/cell_type="Erythroid Precursor Cells"

/cell_line="Primary Culture of Peripheral Blood

Mononuclear Cells"

/dev_stage="Precursor erythroblasts; GPA++"

/lab_host="DH5alpha"

/clone_lib="Hembase; Erythroid Precursor Cells (LCB:cl

library)"

/note="Organ: blood; Vector: pTriplEx2; Site 1: Sfil;

Site 2: Sfil; A complementary DNA (cDNA) library from

human erythroid precursor cells was constructed using

SMART PCR (polymerase chain reaction) cDNA library

Construction Kit (Clontech, Palo Alto, CA) according to

the manufacturer's directions, but with slight

modifications. Briefly, reverse transcription was

performed in the presence of 1 umol/L peptide nucleic acid

(PNA) oligos (N-terminal

)-biotin-GTC-CAC-CCG-AAG-CTT-G-(C-terminal) and

(N-terminal)-biotin-C(T/C)-T-GAA-GTT-CTC-AGG-A-(C-terminal

). Synthesized cDNA was digested with Sfil and

size-selected on a 1% agarose gel (>800bp). Large-scale

sequencing of the library was performed by the NIH

Intramural Sequencing Center (NISC;

Http://www.nisc.nih.gov/."

BASE COUNT 104 a 136 c 150 g 125 t

ORIGIN

Query Match

97.8%; Score 428.4; DB 13; Length 515;

Best Local Similarity 98.6%; Pred. No. 5.3e-103;

Matches 432; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

1 GTGCACCTGACTCTCTGAGAGAGTGGCGGCTTACTGCGCTGTGGGCAAGTGAACGTG 60

Db

63 GTGCACCTGACTCTCTGAGAGAGTGGCGGCTTACTGCGCTGTGGGCAAGTGAACGTG 122

QY

61 GATGAAGTTGGTGGTGGAGCTGGGCGGCTGGTGGTGTCTACCTTGGACCCAGAGG 120

Db

123 GATGAAGTTGGTGGTGGAGCTGGGCGGCTGGTGGTGTCTACCTTGGACCCAGAGG 182

1	428.4	97.8	625	10	US-09-782-378A-14	Sequence 14, Appl
2	428.4	97.8	809	15	US-10-198-846-9489	Sequence 9489, Ap
3	425.2	97.1	438	9	US-09-839-164-3	Sequence 3, Appl
4	425.2	97.1	438	14	US-10-085-853-32	Sequence 32, Appl
5	425.2	97.1	438	15	US-10-128-581-22	Sequence 22, Appl
6	422	96.3	1140	10	US-09-986-666-5	Sequence 5, Appl
7	422	96.3	1140	10	US-09-986-666-7	Sequence 7, Appl
8	422	96.3	1140	10	US-09-986-667-5	Sequence 5, Appl
9	422	96.3	1140	10	US-09-986-667-7	Sequence 7, Appl
10	422	96.3	1140	10	US-09-986-633-5	Sequence 5, Appl
11	422	96.3	1140	10	US-09-986-633-7	Sequence 7, Appl
12	420.4	96.0	743	13	US-10-280-679B-9	Sequence 9, Appl
13	403.4	92.1	453	11	US-09-918-995-36836	Sequence 36836, A
14	393.2	89.8	468	10	US-09-880-107-2305	Sequence 2305, Ap
15	384.4	87.8	626	15	US-10-198-846-9446	Sequence 9446, Ap

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2003, 14:17:18 ; Search time 279 Seconds
(without alignments)
4237.827 Million cell updates/sec

Title: US-09-787-216A-2

Perfect score: 438

Sequence: 1 gtgcacctgactcctgagga.....ccctggccacacagatcatc 438

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Genesec 19Jun03:*

- 1: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/genesec/genesecq/emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/genesec/genesecq/emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/genesec/genesecq/emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/genesec/genesecq/emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/genesec/genesecq/emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/genesec/genesecq/emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	438	100.0	438	21 AAZ94729
2	428.4	97.8	438	22 AAF31396
3	428.4	97.8	625	24 ABE56982
4	428.4	97.8	634	13 AAQ20238
5	425.2	97.1	438	14 AAQ49616
6	425.2	97.1	438	17 AAT29038
7	425.2	97.1	438	18 AAT92697
8	425.2	97.1	626	15 AAQ55625

Human adult beta h
Human alpha and be
Human alpha and be
Encodes human beta
Encodes human beta
Petunia EFSP synth
Beta-haemoglobin c
Human beta-haemogl
DNA sequence encod
Mutant beta-hemogl
Encodes val(1) to
Human breast cance
Human cDNA differe
Gene #2305 used to
Kidney cancer rela
Human polynucleoti
Human cDNA SEQ ID
Human full-length
Human hemoglobin a
Human adult delta
Rabbit beta globin
Rabbit beta-globin
Human secreted pro
Nucleotide sequenc
Rat EX clone nucle
Human polynucleoti
Human contig polyn
(Des-Val)-alpha-(G
Polycistronic oper
Human hemoglobin e
Human transport pr
Human embryonic ep
Polycistronic oper
Human haemoglobin
Human cDNA #262 di

ALIGNMENTS

RESULT 1
AAZ94729
ID AAZ94729 standard; DNA; 438 BP.
XX
AC AAZ94729;
XX
DT 01-AUG-2000 (first entry)
XX
DE DNA encoding human mutant beta-globin.
XX
KW Beta-globin; haemoglobin; Hb Prisca; human; mutant;
KW blood substitute; antianemic; oxygen-carrier; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FN WO200018802-A1.
XX
PD 06-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22756.
XX
PR 01-OCT-1998; 98US-0102640.
XX
(UTEM) UNIV TEMPLE.
XX
PI Fronticelli C;
XX
DR WPI; 2000-293112/25.
DR P-PSDB; AAY79378.
XX

